כותרת אופציונלית: <u>**בינה מלאכותית כללית – הייתכן?**</u>

למקרה שאתם לא יודעים, בעולם כעת מתרחש מירוץ חימוש אחר מי ישיג את הבינה המלאכותית הנחשקת. לא מאמינים? הנה לכם כתבה לדוגמא: https://www.pc.co.il/news/world-news/425210/

וזה לא רק בין סין לארה"ב. זה גם ANTHROPIC שבצרפת. זה איראן, וזה גם איחוד האמירויות.

אבל על איזו בינה מלאכותית מדובר? האם מדובר בכלי שיעזור לכם לכתוב מיילים טובים יותר? אולי ימציא עבורכם סיפור נחמד? יעזור לפתור שיעורי בית או עבודות? או אפילו יעזור לכם בעבודה, כמו בכתיבת או בתיקון של קוד?

ממש אבל ממש לא!

כל אלו כבר די קיימים, אמנם לא מושלמים, ונחשבים יותר כמו עוזרים אישיים. עוזרים לבן אדם בעולמו הפרטי, הצר והקטן, מבודדים אותנו עוד יותר כמו טכנולוגיות אחרות (פלאפונים, אתרי הכרויות, ועוד).

אנחנו מדברים על מירוץ להשגת בינה מלאכותית כללית!

אבל מה זה? ואיך זה שונה מבינה מלאכותית הרגילה הידועה לנו?

הבינה המלאכותית שהיתה ידועה לנו עד שנות ה2023 ועדיין קיימת היא בעצם סט של כלים (שיטות ואלגוריתמים) כלליים, שניתן בעזרתם לפתור בעיות קונקרטיות. כלומר "להלביש" אותם על גבי איזשהו דאטאסט קיים ולעבד אותו לצרכים שלנו. עושים את זה כיום בכל תחום אפשרי: בתחומי חקלאות, מים, שיווק, גיוס עובדים, רפואה, סייבר, בטחון, ועוד. ראינו דוגמאות לזה בכלים פנטסטיים שהתעלו מעל ביצועים אנושיים, כמו בזיהוי אובייקטים בתמונות ובוידאו, במשחקי GO ושחמט, בג'ינרוט חלבונים, ועוד ועוד.

אבל משהו השתנה ב2023, וכולנו חווינו את זה – היציאה של CHATGPT לעולם. לראשונה היה לנו כלי כללי, שמסוגל לענות להרבה מאוד צרכים שונים, ולא רק עונה לבעיות קונקרטיות ולא מבצע אך ורק משימות מסויימות. עם חיזוק תחום הרובוטיקה לאחרונה, כלים אלו, של מודלי שפה, ביחד עם יכולות של סוכנים, מסוגלים לבצע פעולות מעבר לטקסט. כלומר להפעיל תוכנות, לבצע פעולות באינטרנט, ועם רובוטים יוכלו אפילו לבצע מטלות בבית. כל זה באמת פנטסטי, וכפי שהזכרנו לעיל, זה יהיה עוזר אישי נפלא ואכן יקל עוד יותר על חיינו, כמו כל טכנולוגיה קודמת שפגשנו.

אבל היציאה של מודלי שפה לעולם ב2023 עשתה דבר נוסף. לראשונה האנושות נפגשה עם כלי שמגיב בצורה מאוד ברורה, דמויית בן אדם, גמישה ואדפטיבית לבקשות רבות ומגוונות של אנשים. זה בעצם כמו גירסה התחלתית של בינה מלאכותית כללית. אבל מה זה?

אם בינה מלאכותית שהיתה ידועה עד 2023 ועדיין קיימת היא כלים לפתרון בעיות ספציפיות, אז בינה מלאכותית כללית זה כלי אחד גנרי. כמו מח מלאכותי. משהו מאוד אדפטיבי, שלומד תוך כדי תנועה, אבל גם יודע לחדש, מעבר למה שאומן. כלומר מסוגל להכליל, בניגוד לכל הכלים שקיימים היום, כולל מודלי שפה – שנראים לנו מרשימים, אבל בסופו של דבר לכולם יש הכללה מאוד מצומצמת. הם מדהימים אותנו בתשובות שלהם, בגלל שלרוב אנחנו פוגעים בול בסוג השאלות שהם אומנו עליהם. אבל ברגע שאנחנו קצת סוטים מזה, אנחנו פוגשים באגים, טעויות ומה שמכונה אשליות.

מעבר לזה, בינה מלאכותית כללית היא גם קונסיסטנטית ולוגית מהתחלה. בדיוק כמו בן אדם. אם תיקח אפילו ילד בן 6 ותשאל אותו שאלה, הוא יסביר לך או יתן תשובה לפי ההגיון שפיתח עד כה. ואפילו אם התשובה היא שגויה, עדיין היא רציונלית. ואפשר לחקור ולעקוב אחר מהלך החשיבה שלו כדי להגיע לשורש, ולתקן אותו איפה שהוא טעה, ואז הוא יעדכן את הידע שלו בהתאם. ככה בעצם כולנו לומדים בבית הספר. כל פעם מעדכנים את מודלי החשיבה והידע שלנו, מחדדים אותם ומדייקים אותם. אבל הם תמיד קונסיסטנטיים. הם רק משתדרגים עם הזמן.

דבר נוסף הוא שבינה מלאכותית כללית צריכה לחדש, ולשם כך יצירתיות צריך להיות רכיב מרכזי בה. שזה קיים היום במינון די נמוך במודלי הלמידה העמוקה הקיימים היום, בצורות הסתברותיות שונות, כמו בעזרת התפלגויות וחוקי בייס.

ומה הבינה המלאכותית הזו תוכל לעשות בשונה ממה שהיא עושה היום?

אוו... המון!

אז כפי שהבהרנו, ניתן לסכם את כל היישומים של בינה מלאכותית כיום ככלי עזר אישי. בינה מלאכותית כללית לעומת זאת תעזור לכלל האנושות. תקדם אותה בתקווה בצורה רצינית. זה כלי שיוכל לפתור בעיות גלובליות, מדעיות וטכנולוגיות. כגון למצוא פתרונות להתחממות גלובלית, פיתוח מחשב קוונטי וטכנולוגיות קוונטיות אחרות (הצפנה), לקדם את הפיזיקה כמו על ידי הצעת מודל איחוד של פיזיקה קוונטית עם תורת היחסות הכללית. תוכל ללמד כל אדם בהתאם לרמה שלו בכל תחום בצורה מאוד אפקטיבית (למידה בקצב מותאם אישי). תוכל לייצר חומרים חדשים למטרות שונות. בתחום הרפואה: ג'ינרוט איברים מלאכותיים, תרופות חדשות ואפילו שיפור איכות החיים והארכתם. תחום מאוד חשוב נוסף הוא אנרגיה: גם בכדור הארץ, אבל גם מעבר לו. היכולת לרתום אנרגיה בסקאלות שלא הכרנו, ואפילו להשתמש בזה למסע בחלל. שם כבר אפשר יהיה לשלוח רובוטים מצויידים בבינה הזו כדי לחקור מקומות פוטנציאלים ליישוב של בני אדם, ואולי אפילו להפעלת תהליכי הארצה שם (הכנת תנאי מחייה מתאימים לבני אדם).

כל זה נשמע פנטסטי, וכמעט מדע בדיוני, אבל זה רק מה שאנחנו יכולים לדמיין כעת. האם אפשר יותר מזה?

כן! בינה מלאכותית כללית כזאת לא חייבת להיות העתק מושלם של חשיבה אנושית, אלא ניתן לחקור ולפתח מבני חשיבה אחרים. אמנם עדיין הכלים הללו יצטרכו להבין אותנו ולתקשר איתנו, כלומר נצטרך גם להבין אותם, כדי ליישם את התשובות שלהם, אך התהליכים הקוגניטיביים הפנימיים יכולים להיות חדשים ושונים.

כל זה יאפשר ליצור אפילו לוגיקה שלא מוכרת לנו, מה שיאפשר גם ליצור אולי מתמטיקה חדשה, שתאפשר בתורה ליצור מודלים חדשים בכל תחום בעולמנו. כל אלו יאפשרו לפתח טכנולוגיות שהן מעבר להישגנו והבנתנו. אם כי אולי זה מעורר את החשש הידוע מחיסול המין האנושי. אנחנו בדעה שזה כלי שהיה ועדיין יישאר בצורתו הרגילה: כלי ריאקטיבי. כלי ללא רצון משלו, אלא רק מגיב לבקשות של היוזר, שזה האדם.

ולכן לא משנה מה הכלי הזה יעולל או יעשה, בין טוב או רע, זה אך ורק יהיה באחריות האנשים המשתמשים בו. אלו שמלמדים אותו, ואלו שמבקשים ממנו משהו. כמו כל טכנולוגיה שהיתה עד כה, כמו פיזיקה גרעינית, אפשר להשתמש לטובה (אנרגיה גרעינית) או לרעה (פצצת אטום/מימן וכ"ו). הכל תמיד היה ויישאר בידי האדם.

בסדר, נשמע מדהים, אבל איך כל זה קשור אלי?

נכון. אפשר מצד אחד תמיד להמשיך בשגרה. וללכת לפי התהליך הטבעי בחיים: לרכוש השכלה, להתמחות במקצוע מסויים, ולהוציא לפועל את הידע שלי בצורה פרקטית בעולם. כנ"ל גם במחקר בבינה מלאכותית שיש היום. ללמוד את הקורסים הרגילים. להתמחות באספקט אחד או בכמה כאלו צרים, ולהציע רעיונות ומחקרים להרחבת ולפיתוח התחומים הצרים הללו. וככה להתקדם בצעדים קטנטנים בשיפור הקיים. מה שנקרא התקדמות אינקרמנטלית. וזה אכן מה שקיים היום.

אבל! וזו האופציה השנייה. אפשר גם לחשוב בגדול ולחיות בגדול! להרחיב אופקים, ולהכיר הרבה מאוד רעיונות ושיטות בבינה מלאכותית וכל התחומים הרלוונטיים, שיכולים להיות פוטנציאל לפיתוח בינה מלאכותית כללית. תחומים כמו פסיכולוגיה, פילוסופיה, מדעי החברה, ובעיקר מדעי המחשב ומדעי המח.

וזה כבר מסלול חיים אחר לגמרי. זה לא ריצה ל100 מטר, זו ריצת מרתון של אלפי קילומטרים. זה מסע חיים מאוד ארוך, בלי הבטחות לכלום, אלא אולי לתוצרי ביניים.

זה מסע הרפתקה נפלא, כמו מסע של רובינזון קרוזו או כריסטופר קולומבוס אחרי ארצות חדשות, או כמו בסטארטרק שכל פרק מגלים עולמות חדשים. זה לחיות כל הזמן בידיעה שאתה נפגש עם עולמות ידע חדשים.

נשמע מאוד מפרך ובעל סיכוי מאוד נמוך להצליח. למה לי?

יש לפחות שני סיבות למה לעבור את כל זה:

אחת זה שזה תחום הכי בסיסי למחקר, כיוון שאנחנו חוקרים את עצם החשיבה שלנו! זה השורש והמקור לכל החדשנות והיצירתיות של כל בני האדם במשך ההיסטוריה. כל תחום, כל מדע וכל טכנולוגיה נוצרו ופותחו בעזרת המנגנון הזה. זה מחקר אחר הדבר הכי מיסתורי והכי מרתק שקיים בעולם הזה. לחקור את עצם

החשיבה, תפיסת המציאות. מי שמוצא מודל מספיק טוב למנגנון זה, ומצליח לממש אותו בצורה כלשהי – בעצם יוצר "חוקר על", שיכול לפתח ולקדם כל תחום שקיים היום ואולי לייצר תחומים חדשים.

שתיים, זה חוזר לכייף שבעצם עיסוק בתחום המופלא הזה. למה? כי אתה כל הזמן מכיר דברים חדשים ולומד דברים חדשים. אבל מה כייף בזה?

אתה נפגש ברעיונות שונים ומשונים. חלקם ביזאריים, חלקם הזויים. חלקם דווקא מעניינים. וחלקם אפילו קשים מאוד להבנה. יוצא שאפשר למקם את כל הרעיונות הללו על סקלת מורכבות. החל מרעיונות מאוד פשוטים, עד לרעיונות סופר מורכבים. אין בעיה להכיל רעיונות פשוטים או אך קצת מורכבים. הבעיה היא מה לעשות כשאתה נתקל ברעיונות מורכבים? לוותר? אבל אולי זה רעיון חשוב, שיכול להיות בעל פוטנציאל גבוה לפיתוח בינה מלאכותית כללית? מה עושים?

מתמודדים! שוברים את הראש. קוראים כמה פעמים. קוראים מכמה מקורות. לומדים את כל חומר הרקע הנדרש. מתייעצים עם חברים או מומחים. מתייעצים עם מודלי שפה כמו CHATGPT! עושים כל מה שאפשר כדי להבין. ומה קורה אז?

או אז אנחנו חוזרים להרפתקאה המופלאה שלנו. ברגע שהצלחנו סופסוף לתפוס את הרעיון ולהבין אותו, זה כמו להשתלט על עולם חדש. זו הרגשה עילאית! הרגשה כל כך טובה, שאחרי מאמץ רב והתמדה ואפילו תסכול, הצלחנו להשתלט על הרעיון, זה פשוט טעם גן עדן. וזה ממכר כמו סם. אתה רק רוצה עוד ועוד מזה.

אז יוצא שאתה רק מחזק את המוטיבציה שלך להתמודד עם רעיונות מסובכים ומורכבים חדשים. זה יוצר חוויות לכל החיים. וזה שווה את המאמץ, גם אם שוב, לא אתה תהיה זה שתבנה את הבינה הזו, אלא אולי מישהו אחר בעתיד. אבל המאמץ שלך הוא לא לשווא. כי כל סיכום ברור של רעיונות שכבשת, כל הסבר שלך לחוקרים אחרים. כל רעיון או תאוריה חדשה שתציע. כולם יכולים לשמש חוקרים אחרים, שאולי אחד מהם כן יצליח איפה שאתה לא הצלחת. זה כמו העברת השרביט בריצת קבוצות. אנחנו כל הזמן מעבירים את ההישגים שלנו, גם אם חלקיים, לרצים הבאים.

ככה המדע והטכנולוגיה התפתחו עד היום. וככה בעזרת ה' נגיע גם לבינה מלאכותית כללית.

זה לא מתאים לכל אחד, ובעצם זה רלוונטי רק לקומץ של אנשים, שמוכנים להקריב ולהקדיש את כל חייהם במרדף אחרי מטרה זו ורק זו. לכן החברה הגלובלית לא צריכה להשקיע הרבה, אלא ממש מעט. סיכון נמוך בהתחשב שגם אם לא יצליח אז הפסדנו קצת. אבל אם מתישהו יצליח, אז זה משהו שיכול להפוך את העולם מקצה לקצה, לחיים שאפילו לא חלמנו עליהם.

בכל מקרה, כפי שהוזכר בתחילת הכתבה: כבר יש כמה חברות גדולות שמתעסקות בזה, וכמה מכונים בודדים בעולם. אז זה כלל לא מופרך, במיוחד אחרי שהכרנו את מודלי השפה המדהימים.

השאלה אם את או אתה רוצים להצטרף למסע?

צרו איתי קשר ואשמח לשמוע מכם. כל עצה וכל דעה מתקבלת בברכה.

shimon-k.github.io

Artificial General Intelligence – Is It Possible?

In case you didn't know, there is currently an arms race to achieve the so desired Artificial Intelligence.

And it's not just between China and the U.S. It's also Anthropic in France. It's Iran, and it's also the United Arab Emirates.

But what kind of Artificial Intelligence are we talking about? Are we referring to a tool that will help you write better emails? Maybe it will invent a nice story for you? Help with homework or tasks? Or even assist you at work, such as generating or correcting code?

Absolutely not!

All of these already exist up to some extent, even if they are not perfect. They are considered more as personal assistants—helping individuals in their private, narrow lives and isolating us further, just like other technologies did so far (phones, dating apps, etc.).

We are talking about the race to achieve **Artificial General Intelligence (AGI)!**

But what is AGI? And how is it different from the Artificial Intelligence we know?

The Difference Between AI and AGI

The Artificial Intelligence known to us until 2023—and still in use—is essentially a set of tools (methods and algorithms) that can be used to solve specific problems. In other words, they are applied to a dataset and processed according to our needs. This is done today in virtually every field: agriculture, water management, marketing, recruitment, medicine, cybersecurity, security, and more. We have seen fantastic tools surpass human performance, such as object recognition in images and video, playing **Go** and chess, protein folding, and many others.

But something changed in 2023, and we all experienced it—the release of **ChatGPT**. For the first time, we had a general tool capable of addressing many different needs, rather than just specific problems or predefined tasks. With recent advances in robotics, these language models, combined with agent-based capabilities, can now perform actions beyond text processing. They can operate software, execute tasks online, and, when paired with robots, even perform household chores and work in factories.

All of this is truly fantastic. As mentioned, these tools will be excellent personal assistants and will ease our lives, much like every other technology we've encountered.

However, the introduction of language models in 2023 did something else. For the first time, humanity encountered a tool that **responds in a clear, human-like, flexible, and adaptive manner** to a wide range of requests. This is essentially a **primitive version of Artificial General Intelligence**.

What is AGI?

If traditional AI tools solve specific problems, then **AGI is a single, universal or generic tool**—like an artificial brain. It is something highly **adaptive**, capable of **learning in real-time**, and even **creating new knowledge** beyond its training data.

Unlike today's models, which have limited generalization capabilities, **AGI can truly generalize**. Current models impress us with their responses because they are well-trained for many of the

questions we ask. But as soon as we deviate slightly from their training data, we encounter **bugs, errors, and hallucinations**.

Additionally, **AGI must be consistent and logical** from the start—just like a human. If you ask a six-year-old a question, he will explain it according to the logic he has developed. Even if his answer is incorrect, it will still be rational. You can follow his thought process, correct him, and he will **update his knowledge accordingly**.

That is how we all learn in school—we constantly **refine and improve** our models of thinking and knowledge. But they are always **consistent** and only improve over time.

Another crucial aspect of **AGI is creativity**. Creativity must be a central component, but in today's deep learning models, it exists only at a **very low level**—mainly through probabilistic methods, such as **distributions and Bayesian inference**.

What Will AGI Be Capable Of?

Oh... a lot!

As we explained, today's AI mainly serves as **a personal assistant**. **AGI, on the other hand, will assist all of humanity.**

It could advance humanity **in unimaginable ways**, solving **global, scientific, and technological challenges**. For example:

- **Finding solutions to climate change**
- **Developing quantum computers and other quantum technologies (encryption)**
- **Advancing physics, such as proposing a unifying model between quantum mechanics and general relativity**
- **Teaching people at their own pace, in any field, in a highly effective manner (personalized learning)**
- **Creating new materials for various applications**
- **Generating artificial organs, developing new medicines, and even improving life expectancy and quality**
- **Harnessing energy at scales beyond our current understanding, even for space travel**

AGI-equipped robots could be sent to explore new planets, preparing them for **human colonization** and even **terraforming** them to create livable conditions.

All of this sounds fantastic, almost like science fiction, but **this is just what we can imagine right now**.

Could there be even **greater possibilities**?

A New Way of Thinking

Yes! **AGI does not have to replicate human thinking exactly**. We can explore and develop **entirely new cognitive structures**.

These systems will still need to **understand and communicate with us**, but their **internal thought processes could be completely different**.

This could lead to **entirely new forms of logic and even new branches of mathematics**, which in turn would enable the development of **technologies beyond our current reach and comprehension**.

Of course, some fear **AGI might lead to human extinction**. But we believe AGI will remain a **reactive tool**, responding only to human requests rather than acting of its own volition.

Thus, no matter what AGI does—whether for good or bad—**humans will always be responsible**. Just like nuclear physics, which can be used for **clean energy** or **destruction (atomic bombs)**, **technology remains in human hands**.

How Does This Concern You?

Fair question. You could continue with life as usual—pursuing education, specializing in a profession, and applying your knowledge practically.

Or...

You could **think big** and **live big!**

The Journey to AGI

Becoming involved in AGI development is **a lifelong journey**. It is **not a sprint—it is a marathon** spanning thousands of miles. It is an **endless adventure**, like **Robinson Crusoe exploring new lands** or **Christopher Columbus seeking unknown territories**.

It is a **path of discovery**—constantly encountering **new worlds of knowledge**.

Sounds exhausting? Maybe. But here's the real reason **why it's worth it**:

1. **This is the most fundamental research topic in existence!**
   We are studying **the very nature of human thought**—the root of all innovation and creativity throughout history. If someone cracks the secret to artificial cognition, they will create **a "super-researcher" capable of advancing all fields**.

2. **The joy of discovery!**
   Encountering new ideas—some bizarre, some crazy, some brilliant—is an adventure. And when you finally understand a difficult concept, it feels **like conquering a new world**.

The struggle to understand **complex ideas** is what makes this journey so **exciting**.

Even if **you are not the one to build AGI**, your contributions—your theories, insights, and discoveries—could **help someone else** succeed in the future.

Science and technology have **always** evolved this way—one step at a time.

So the question is:

Do you want to join the journey?

Contact me—I'd love to hear your thoughts!

shimon-k.github.io